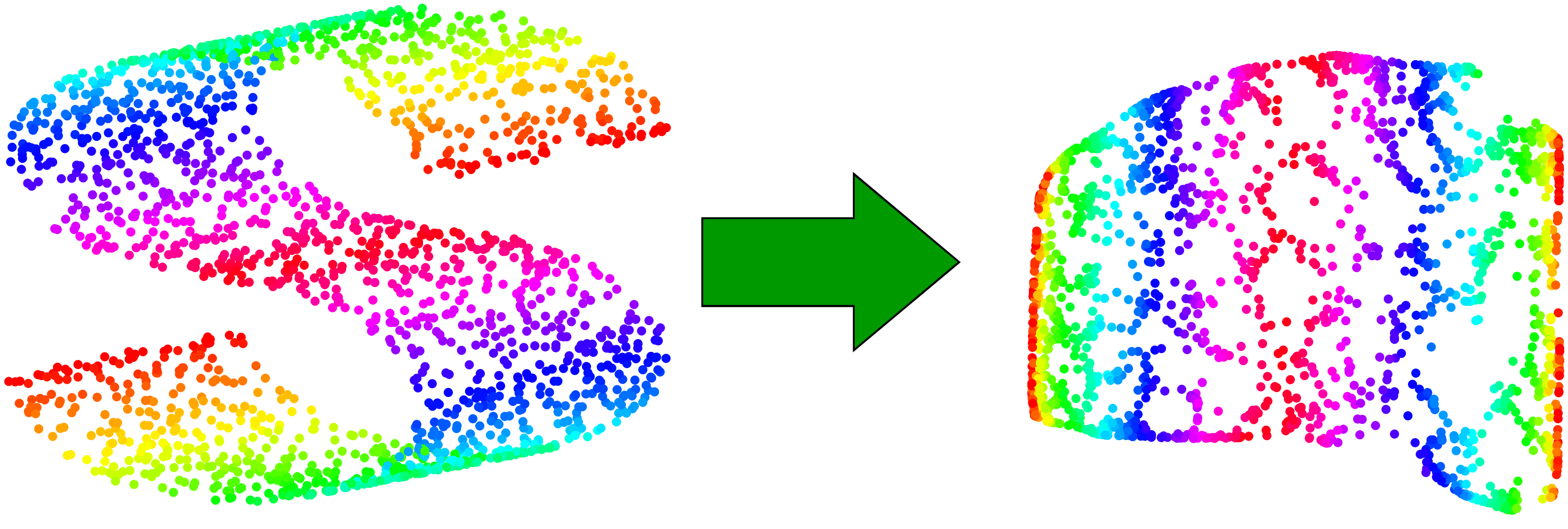
教師付き線形次元削減（１７章）

杉山将・本多淳也
sugi@k.u-tokyo.ac.jp, jhonda@k.u-tokyo.ac.jp
http://www.ms.k.u-tokyo.ac.jp

次元の呪い

入力の次元dが増えると
学習問題は指数的に難しくなる

■本質的な情報を保持したまま次元を減らしたい！



線形次元削減

- （高次元）標本：

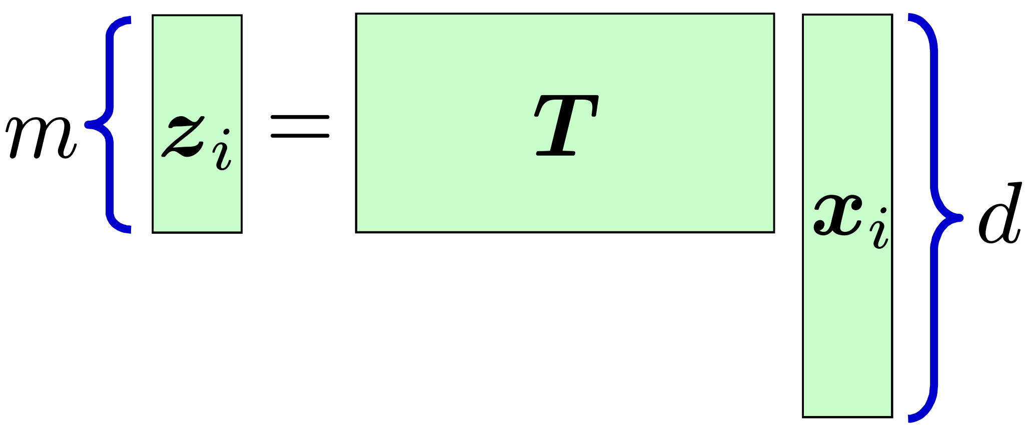
$$\{\boldsymbol{x}_i\}_{i=1}^n, \quad \boldsymbol{x}_i \in \mathbb{R}^d, \quad d \gg 1$$

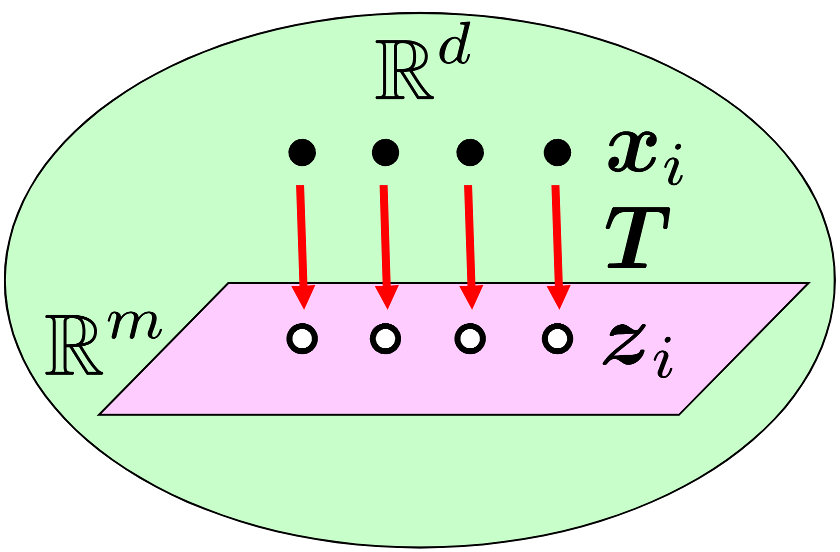
- 埋め込み行列：

$$\boldsymbol{T} \in \mathbb{R}^{m \times d}, \quad 1 \leq m \ll d$$

- 埋め込まれた標本

$$\{\boldsymbol{z}_i\}_{i=1}^n, \quad \boldsymbol{z}_i = \boldsymbol{T}\boldsymbol{x}_i \in \mathbb{R}^m$$

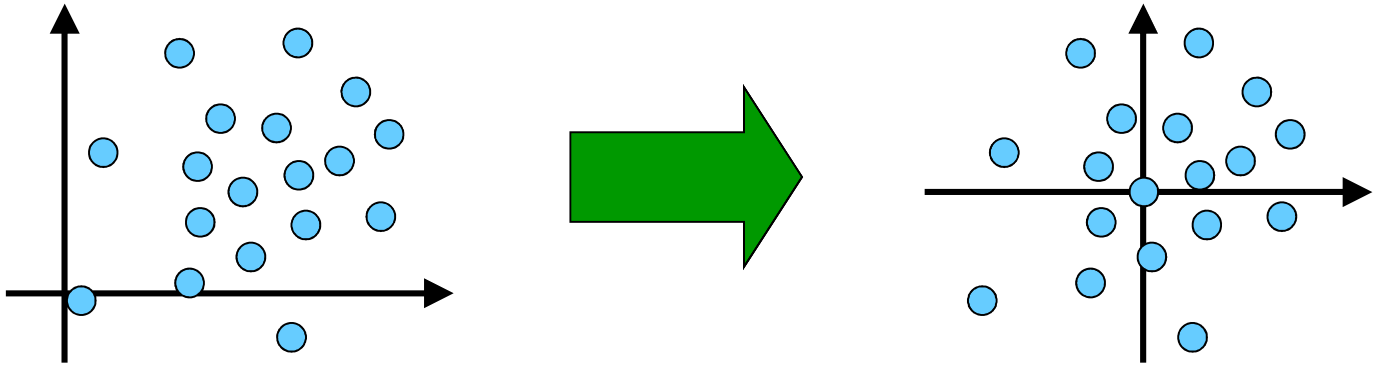




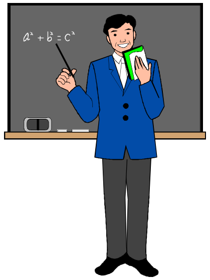
標本の中心化

- 以後，簡便のため標本を中心化して平均がゼロになるようにしておく

$$\boldsymbol{x}_i \longleftarrow \boldsymbol{x}_i - \frac{1}{n}\sum_{i'=1}^{n} \boldsymbol{x}_{i'}$$



$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i = 0$$



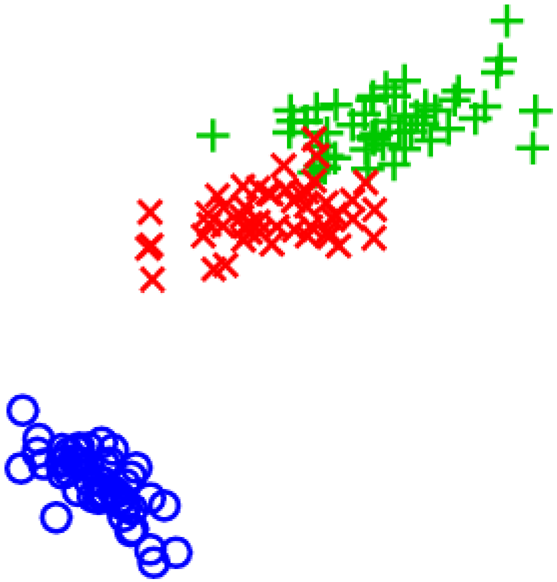
講義の流れ

1. フィッシャー判別分析（19章）
2. 局所フィッシャー判別分析（19章）
3. 半教師付き局所フィッシャー判別分析（19章）
4. 十分次元削減（19章）

教師付き次元削減

- クラスラベル付きの標本: $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$
  $\boldsymbol{x}_i \in \mathbb{R}^d$
  $y_i \in \{1, 2, \ldots, c\}$

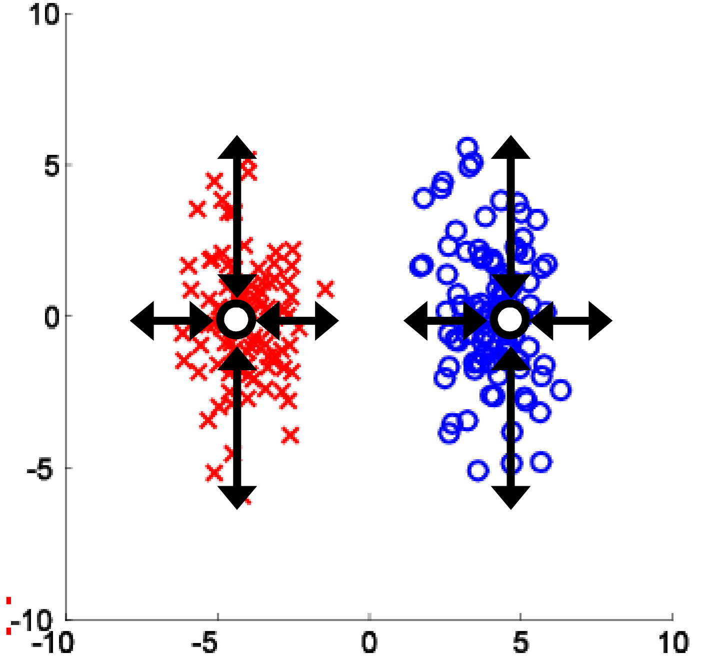
- 異なるクラスの標本を分けるように次元削減を行ないたい.



散布行列

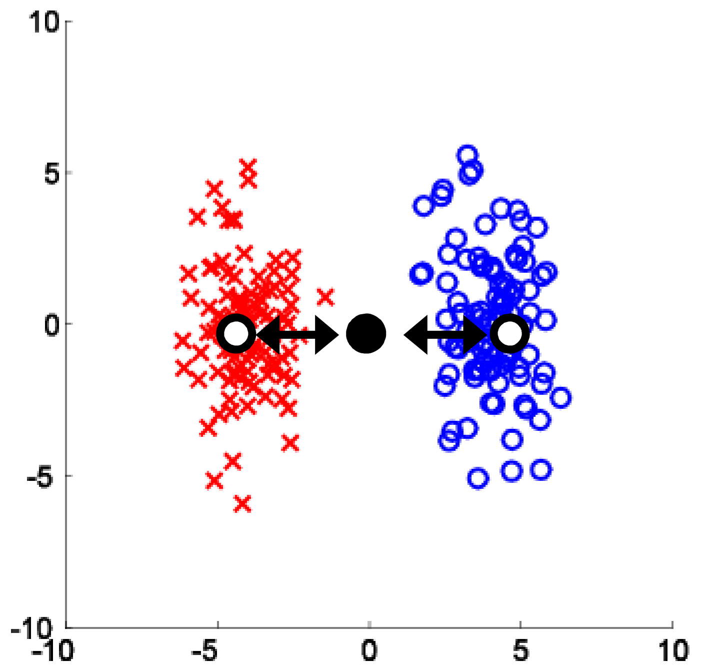
■ クラス内(within-class)散布行列：

$$\boldsymbol{S}^{(\mathrm{w})} = \sum_{y=1}^{c} \sum_{i:y_i=y} (\boldsymbol{x}_i - \boldsymbol{\mu}_y)(\boldsymbol{x}_i - \boldsymbol{\mu}_y)^\top$$



■ クラス間(between-class)散布行列：

$$\boldsymbol{S}^{(\mathrm{b})} = \sum_{y=1}^{c} n_y \boldsymbol{\mu}_y \boldsymbol{\mu}_y^\top$$



n_y :クラス y の標本数

$$\boldsymbol{\mu}_y = \frac{1}{n_y} \sum_{i:y_i=y} \boldsymbol{x}_i \qquad \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i = 0$$

散布行列の関係

- 散布行列:
$$\boldsymbol{C} = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top$$
- クラス間散布行列:
$$\boldsymbol{S}^{(\mathrm{b})} = \sum_{y=1}^{c} n_y \boldsymbol{\mu}_y \boldsymbol{\mu}_y^\top$$
- クラス内散布行列:
$$\boldsymbol{S}^{(\mathrm{w})} = \sum_{y=1}^{c} \sum_{i:y_i=y} (\boldsymbol{x}_i - \boldsymbol{\mu}_y)(\boldsymbol{x}_i - \boldsymbol{\mu}_y)^\top$$
- 数学演習:以下の関係を示せ

$$\boldsymbol{C} = \boldsymbol{S}^{(\mathrm{w})} + \boldsymbol{S}^{(\mathrm{b})}$$

解答例

$$\boldsymbol{S}^{(\mathrm{w})} = \sum_{y=1}^{c} \sum_{i:y_i=y} \left(\boldsymbol{x}_i \boldsymbol{x}_i^\top - 2\boldsymbol{x}_i \boldsymbol{\mu}_y^\top + \boldsymbol{\mu}_y \boldsymbol{\mu}_y^\top\right)$$

$$= \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top - 2\sum_{y=1}^{c} n_y \boldsymbol{\mu}_y \boldsymbol{\mu}_y^\top + \sum_{y=1}^{c} n_y \boldsymbol{\mu}_y \boldsymbol{\mu}_y^\top = \boldsymbol{C} - \boldsymbol{S}^{(\mathrm{b})}$$

フィッシャー判別分析 (FDA: Fisher Discriminant Analysis)

- 埋め込み後の標本 $\boldsymbol{z}_i = \boldsymbol{T}\boldsymbol{x}_i$ のクラス内・クラス外散布行列はそれぞれ $\boldsymbol{T}\boldsymbol{S}^{(\mathrm{b})}\boldsymbol{T}^\top, \boldsymbol{T}\boldsymbol{S}^{(\mathrm{w})}\boldsymbol{T}^\top$

- クラス内散布を小さく，クラス間散布を大きくする：

$$\boldsymbol{T}_{\mathrm{FDA}} = \underset{\boldsymbol{T}\in\mathbb{R}^{m\times d}}{\mathrm{argmax}}\,\mathrm{tr}\left((\boldsymbol{T}\boldsymbol{S}^{(\mathrm{w})}\boldsymbol{T}^\top)^{-1}\boldsymbol{T}\boldsymbol{S}^{(\mathrm{b})}\boldsymbol{T}^\top\right)$$

フィッシャー判別分析 (FDA: Fisher Discriminant Analysis)

- クラス内散布を小さく，クラス間散布を大きくする:

$$\boldsymbol{T}_{\mathrm{FDA}} = \mathop{\mathrm{argmax}}_{\boldsymbol{T}\in\mathbb{R}^{m\times d}} \mathrm{tr}\left((\boldsymbol{T}\boldsymbol{S}^{(\mathrm{w})}\boldsymbol{T}^\top)^{-1}\boldsymbol{T}\boldsymbol{S}^{(\mathrm{b})}\boldsymbol{T}^\top \right)$$

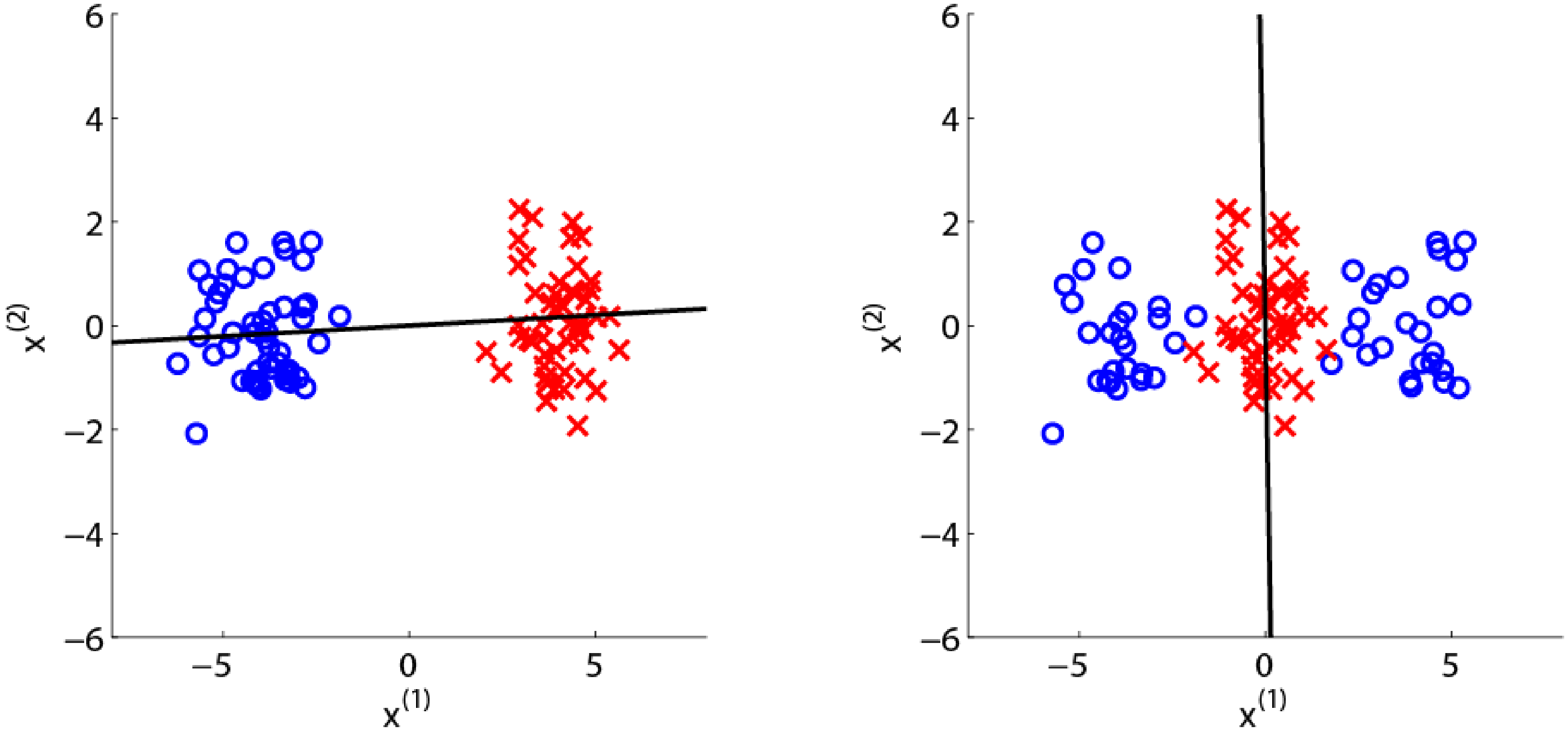
- 解の求め方:
  - 一般化固有値問題を解く: $\boldsymbol{S}^{(\mathrm{b})}\boldsymbol{\xi} = \lambda \boldsymbol{S}^{(\mathrm{w})}\boldsymbol{\xi}$

$$\lambda_1 \geq \cdots \geq \lambda_d \qquad \boldsymbol{\xi}_j^\top \boldsymbol{S}^{(\mathrm{w})}\boldsymbol{\xi}_j = 1$$

  - 上位 m 個の固有ベクトルを並べる

$$\boldsymbol{T}_{\mathrm{FDA}} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)^\top$$

実行例



- フィッシャー判別分析によって，異なるクラスのデータを分離する部分空間が得られている
- しかし，クラス内にクラスタ構造があるとうまくいかないことがある

実装は宿題

フィッシャー判別分析の解

- $y_i = \pm 1$ の2クラス分類問題を考え，フィッシャー判別分析により $m = 1$ 次元の部分空間を見つける

$$\boldsymbol{t}_{\mathrm{FDA}} = \underset{\boldsymbol{t} \in \mathbb{R}^d}{\mathrm{argmax}} \frac{\boldsymbol{t}^\top \boldsymbol{S}^{(\mathrm{b})} \boldsymbol{t}}{\boldsymbol{t}^\top \boldsymbol{S}^{(\mathrm{w})} \boldsymbol{t}}$$

- 目的関数がスケール変換に対して不変なのでこの最適化問題は以下と等価：

$$\boldsymbol{t}_{\mathrm{FDA}} = \underset{\boldsymbol{t} \in \mathbb{R}^d}{\mathrm{argmax}}\, \boldsymbol{t}^\top \boldsymbol{S}^{(\mathrm{b})} \boldsymbol{t} \quad \text{subject to } \boldsymbol{t}^\top \boldsymbol{S}^{(\mathrm{w})} \boldsymbol{t} = 1$$

- ラグランジュの乗数法より

$$\boldsymbol{S}^{(\mathrm{b})} \boldsymbol{t}_{\mathrm{FDA}} + \lambda \boldsymbol{S}^{(\mathrm{w})} \boldsymbol{t}_{\mathrm{FDA}} = 0$$

数学演習

$$\boldsymbol{S}^{(\mathrm{b})}\boldsymbol{t}_{\mathrm{FDA}} + \lambda \boldsymbol{S}^{(\mathrm{w})}\boldsymbol{t}_{\mathrm{FDA}} = 0$$

- フィッシャー判別分析の解が以下の形で表されることを示せ：

$$\boldsymbol{t}_{\mathrm{FDA}} \propto \boldsymbol{C}^{-1}\boldsymbol{\mu}_{+1}$$

- ヒント：
  - $\boldsymbol{C} = \boldsymbol{S}^{(\mathrm{w})} + \boldsymbol{S}^{(\mathrm{b})}$
  - $\boldsymbol{S}^{(\mathrm{b})} = n_{+1}\boldsymbol{\mu}_{+1}\boldsymbol{\mu}_{+1}^{\top} + n_{-1}\boldsymbol{\mu}_{-1}\boldsymbol{\mu}_{-1}^{\top}$
  - 中心化の仮定より $n_{+1}\boldsymbol{\mu}_{+1} + n_{-1}\boldsymbol{\mu}_{-1} = \boldsymbol{0}$
  - 任意の t に対し $(\boldsymbol{\mu}_{+1}\boldsymbol{\mu}_{+1}^{\top})\boldsymbol{t} = \boldsymbol{\mu}_{+1}(\boldsymbol{\mu}_{+1}^{\top}\boldsymbol{t}) \propto \boldsymbol{\mu}_{+1}$

解答例

- $\boldsymbol{S}^{(\mathrm{b})}\boldsymbol{t}_{\mathrm{FDA}} + \lambda \boldsymbol{S}^{(\mathrm{w})}\boldsymbol{t}_{\mathrm{FDA}} = 0$と$\boldsymbol{C} = \boldsymbol{S}^{(\mathrm{w})} + \boldsymbol{S}^{(\mathrm{b})}$より

$$\boldsymbol{S}^{(\mathrm{b})}\boldsymbol{t}_{\mathrm{FDA}} + \lambda(\boldsymbol{C} - \boldsymbol{S}^{(\mathrm{b})})\boldsymbol{t}_{\mathrm{FDA}} = 0$$

- C^{-1}を左からかけて整理すると

$$\lambda \boldsymbol{t}_{\mathrm{FDA}} = (\lambda - 1)\boldsymbol{C}^{-1}\boldsymbol{S}^{(\mathrm{b})}\boldsymbol{t}_{\mathrm{FDA}}$$

- 一方で

$$\boldsymbol{S}^{(\mathrm{b})} = n_{+1}\boldsymbol{\mu}_{+1}\boldsymbol{\mu}_{+1}^{\top} + n_{-1}\boldsymbol{\mu}_{-1}\boldsymbol{\mu}_{-1}^{\top}$$

$$= n_{+1}\boldsymbol{\mu}_{+1}\boldsymbol{\mu}_{+1}^{\top} + \frac{n_{-1}^2}{n_{-1}}\boldsymbol{\mu}_{-1}\boldsymbol{\mu}_{-1}^{\top} \propto \boldsymbol{\mu}_{+1}\boldsymbol{\mu}_{+1}^{\top}$$

なので

$$\boldsymbol{t}_{\mathrm{FDA}} \propto \boldsymbol{C}^{-1}(\boldsymbol{\mu}_{+1}\boldsymbol{\mu}_{+1}^{\top})\boldsymbol{t}_{\mathrm{FDA}} \propto \boldsymbol{C}^{-1}\boldsymbol{\mu}_{+1}$$

フィッシャー判別分析の等価性

ガウス生成モデルの最尤推定での分離平面:

$$\boldsymbol{a}^\top \boldsymbol{x} + b = 0 \qquad \boldsymbol{a} = \boldsymbol{C}^{-1}\boldsymbol{\mu}_{+1}$$

線形最小二乗分類:

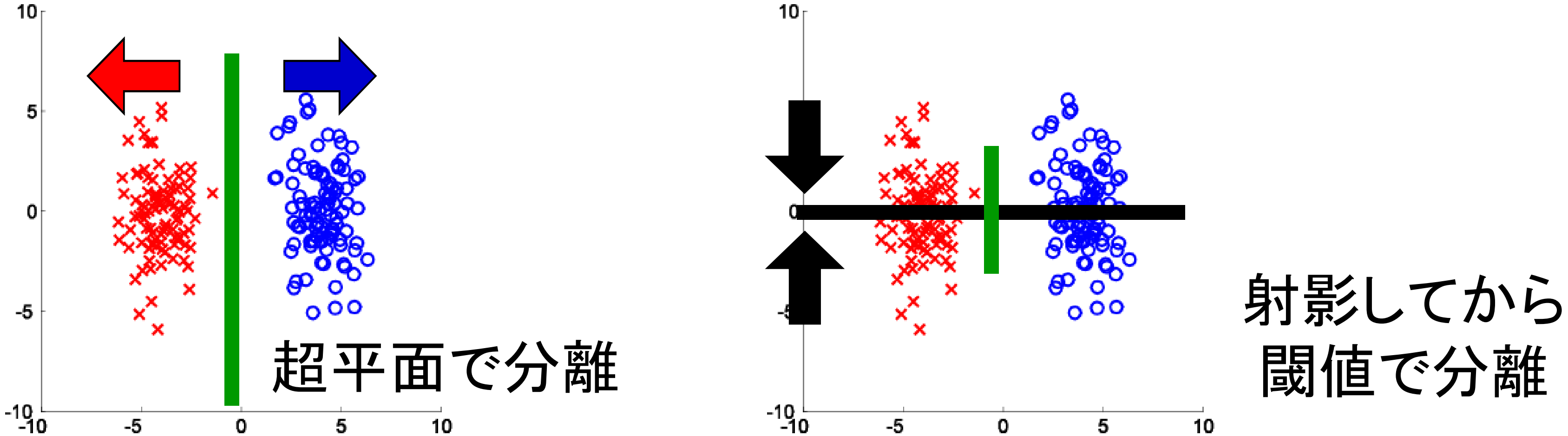
$$y_i \in \left\{\frac{n}{n_{+1}}, \frac{n}{n_{-1}}\right\}$$

$$\widehat{\boldsymbol{\theta}} = \mathrm{argmin}_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left(\boldsymbol{\theta}^\top \boldsymbol{x}_i - y_i\right)^2 \propto \boldsymbol{C}^{-1}\boldsymbol{\mu}_{+1}$$

線形次元削減法:

$$\boldsymbol{t}_{\mathrm{FDA}} \propto \boldsymbol{C}^{-1}\boldsymbol{\mu}_{+1}$$

- 1次元部分空間に射影した後にしきい値処理



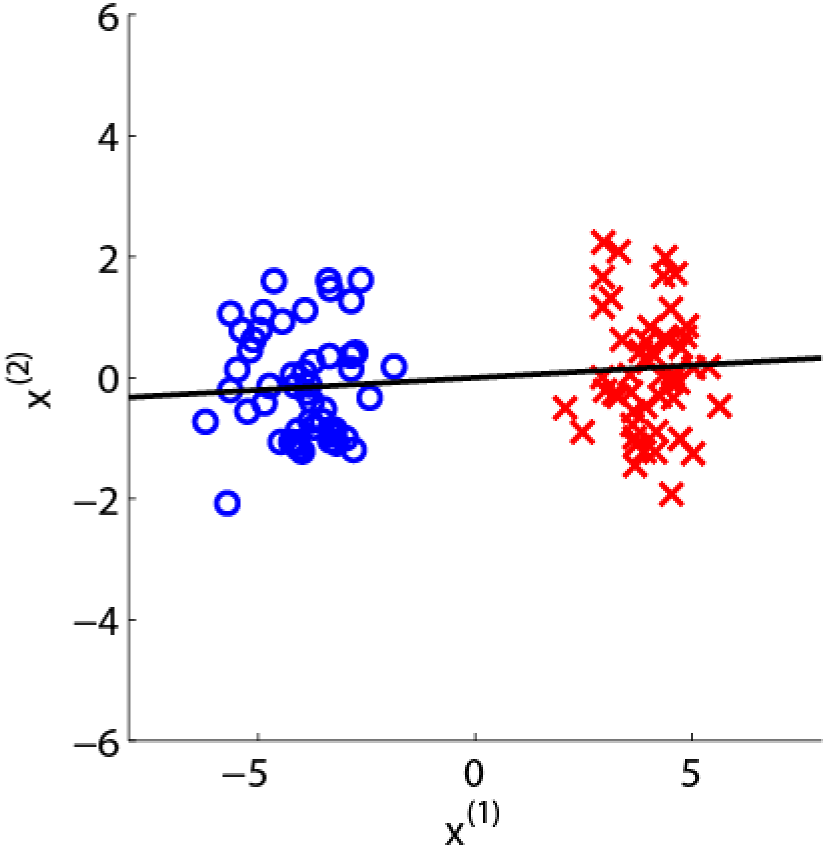
埋め込み次元数の上限

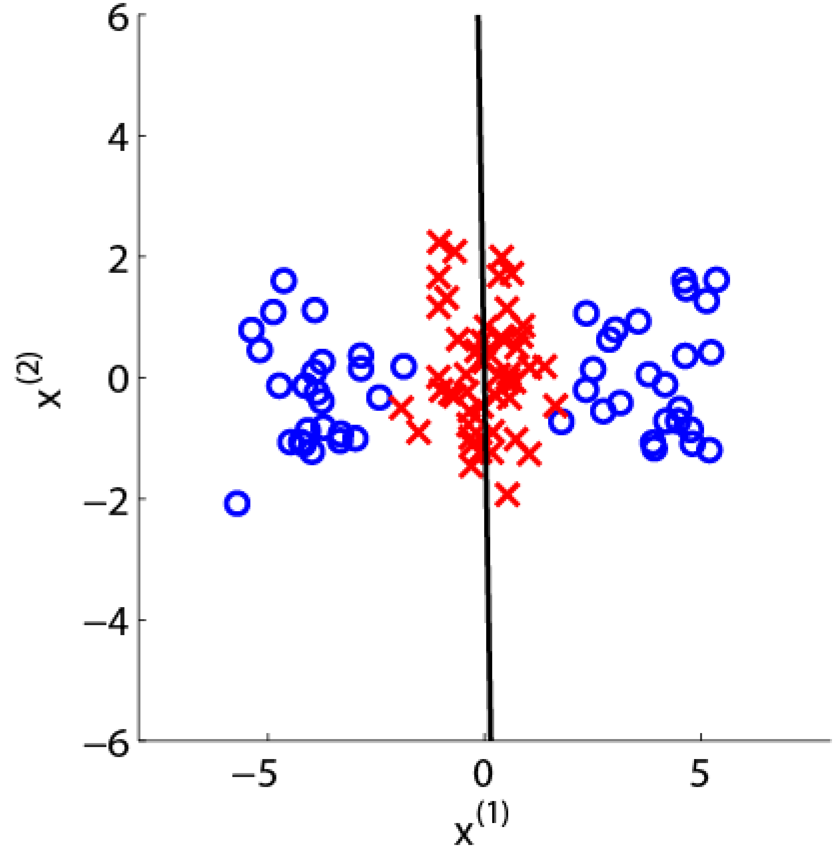
- $\mathrm{rank}(\boldsymbol{S}^{(\mathrm{b})}) \leq c-1$が成り立つ　c :クラス数
  - 証明: $\boldsymbol{S}^{(\mathrm{b})}$のランク＝空間$\{\boldsymbol{S}^{(\mathrm{b})}\boldsymbol{t} : \boldsymbol{t} \in \mathbb{R}^d\}$の次元

$$\boldsymbol{S}^{(\mathrm{b})}\boldsymbol{t} = \sum_{y=1}^{c} n_y \boldsymbol{\mu}_y \boldsymbol{\mu}_y^\top \boldsymbol{t} = \sum_{y=1}^{c} n_y (\boldsymbol{\mu}_y^\top \boldsymbol{t}) \cdot \boldsymbol{\mu}_y$$

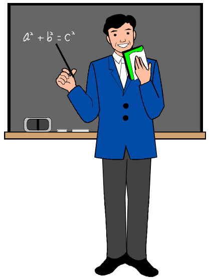
  - 中心化の仮定 $\displaystyle\sum_{y=1}^{c} n_y \boldsymbol{\mu}_y = \boldsymbol{0}$より$\{\boldsymbol{\mu}_y\}$は一時従属
- 一般化固有値問題 $\boldsymbol{S}^{(\mathrm{b})}\boldsymbol{\xi} = \lambda \boldsymbol{S}^{(\mathrm{w})}\boldsymbol{\xi}$の固有値ゼロに対応する解は無意味
- 埋め込み空間の次元数mは最大 $c-1$
  - $c=2$のときは高々$m=1$となってしまう

フィッシャー判別分析：まとめ





- 一般化固有ベクトルから解析的に解が求まる
- 異なるクラスのデータを分離できる
- クラス内クラスタ構造があるとうまくいかないことがある
- 埋め込み空間の次元に上限がある



講義の流れ

1. フィッシャー判別分析（19章）
2. 局所フィッシャー判別分析（19章）
3. 半教師付き局所フィッシャー判別分析（19章）
4. 十分次元削減（19章）

散布行列のペア表現

- クラス内:

$$\boldsymbol{S}^{(\mathrm{w})} = \frac{1}{2}\sum_{i,i'=1}^{n} Q_{i,i'}^{(\mathrm{w})}(\boldsymbol{x}_i - \boldsymbol{x}_{i'})(\boldsymbol{x}_i - \boldsymbol{x}_{i'})^\top$$

$$Q_{i,i'}^{(\mathrm{w})} = \begin{cases} 1/n_y & (y_i = y_{i'} = y) \\ 0 & (y_i \neq y_{i'}) \end{cases}$$

- クラス間:

$$\boldsymbol{S}^{(\mathrm{b})} = \frac{1}{2}\sum_{i,i'=1}^{n} Q_{i,i'}^{(\mathrm{b})}(\boldsymbol{x}_i - \boldsymbol{x}_{i'})(\boldsymbol{x}_i - \boldsymbol{x}_{i'})^\top$$

導出は宿題

$$Q_{i,i'}^{(\mathrm{b})} = \begin{cases} 1/n - 1/n_y \leq 0 & (y_i = y_{i'} = y) \\ 1/n & (y_i \neq y_{i'}) \end{cases}$$

n : 全標本数　　n_y : クラス y の標本数

射影後の散布行列

- クラス内: $\boldsymbol{T}\boldsymbol{S}^{(\mathrm{w})}\boldsymbol{T}^\top = \frac{1}{2}\sum_{i,i'=1}^{n} Q_{i,i'}^{(\mathrm{w})}(\boldsymbol{z}_i - \boldsymbol{z}_{i'})(\boldsymbol{z}_i - \boldsymbol{z}_{i'})^\top$

$$Q_{i,i'}^{(\mathrm{w})} = \begin{cases} 1/n_y & (y_i = y_{i'} = y) \\ 0 & (y_i \neq y_{i'}) \end{cases}$$

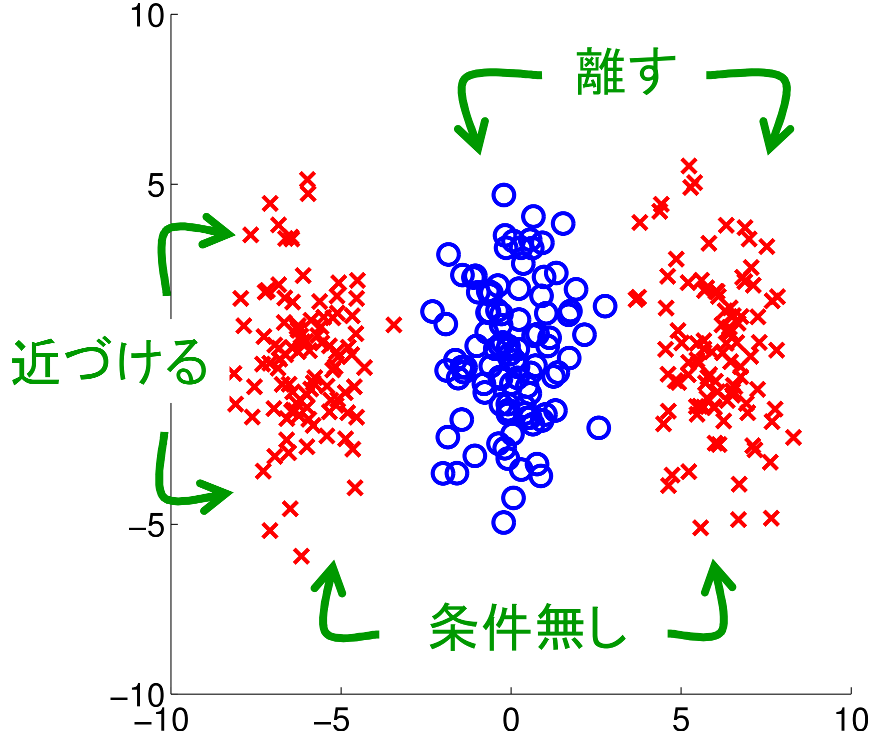
- クラス間: $\boldsymbol{T}\boldsymbol{S}^{(\mathrm{b})}\boldsymbol{T}^\top = \frac{1}{2}\sum_{i,i'=1}^{n} Q_{i,i'}^{(\mathrm{b})}(\boldsymbol{z}_i - \boldsymbol{z}_{i'})(\boldsymbol{z}_i - \boldsymbol{z}_{i'})^\top$

$$Q_{i,i'}^{(\mathrm{b})} = \begin{cases} 1/n - 1/n_y \leq 0 & (y_i = y_{i'} = y) \\ 1/n & (y_i \neq y_{i'}) \end{cases}$$

- FDA: $\boldsymbol{T}\boldsymbol{S}^{(\mathrm{w})}\boldsymbol{T}^\top$は小さく, $\boldsymbol{T}\boldsymbol{S}^{(\mathrm{b})}\boldsymbol{T}^\top$は大きくする
  - 同じクラスの標本は近づける
  - 異なるクラスの標本は遠ざける

局所フィッシャー判別分析 (LFDA: Local FDA)

- 考え方：データの局所性も考慮に入れる
  - A) 同じクラスの近くの標本は近づける
  - B) 異なるクラスの標本は遠ざける
  - C) 同じクラスの遠くの標本は近づける必要はない



局所フィッシャー判別分析
||
フィッシャー判別分析
＋
局所性保存射影

局所フィッシャー判別分析の規準[23]

$\boldsymbol{W}$: 類似度行列

- 局所クラス内散布行列:

$$\boldsymbol{S}^{(\mathrm{lw})} = \frac{1}{2}\sum_{i,i'=1}^{n} Q_{i,i'}^{(\mathrm{lw})}(\boldsymbol{x}_i - \boldsymbol{x}_{i'})(\boldsymbol{x}_i - \boldsymbol{x}_{i'})^\top$$

$$Q_{i,i'}^{(\mathrm{lw})} = \begin{cases} W_{i,i'}/n_y & (y_i = y_{i'} = y) \\ 0 & (y_i \neq y_{i'}) \end{cases}$$

- 局所クラス間散布行列:

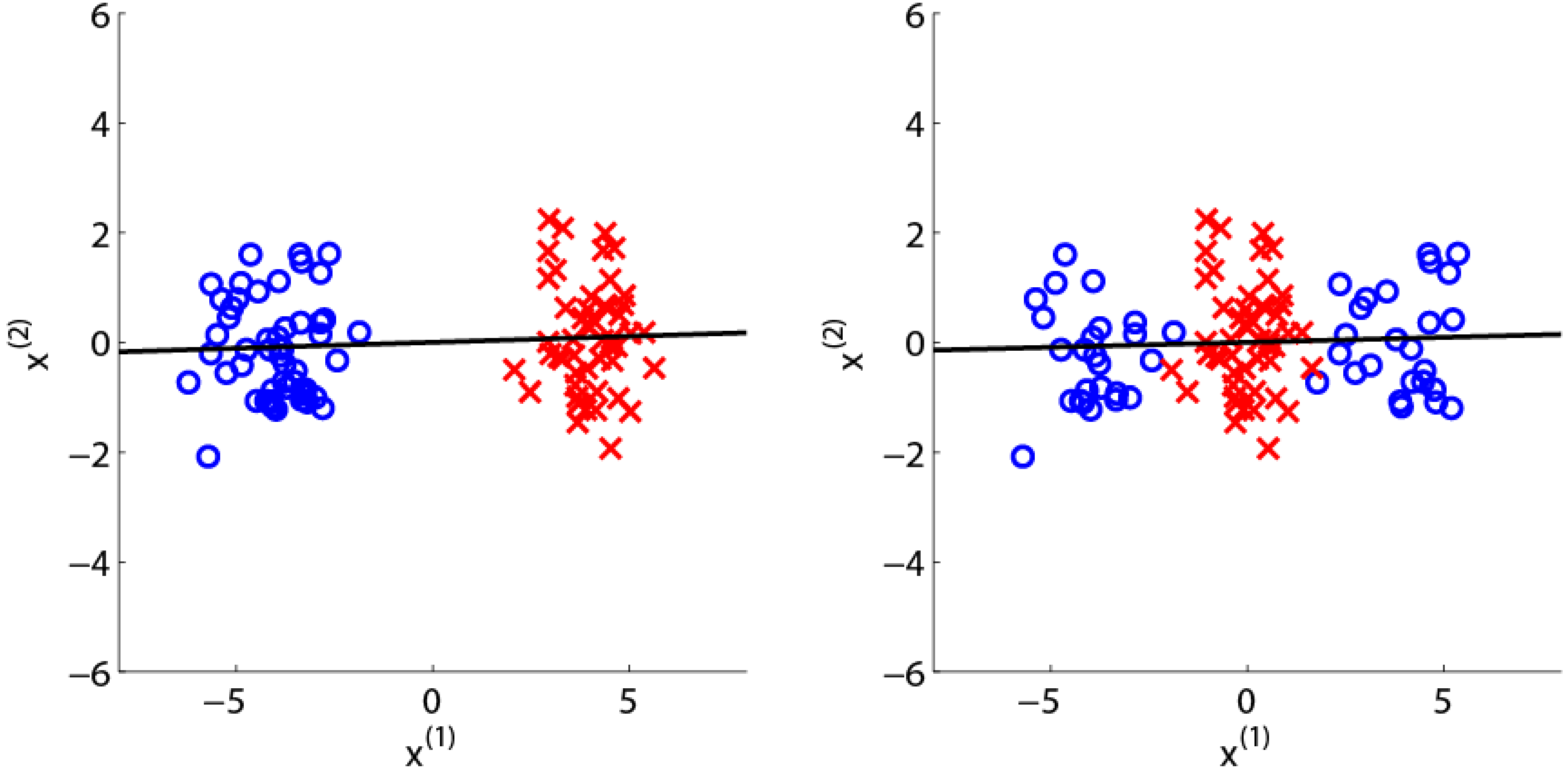
$$\boldsymbol{S}^{(\mathrm{lb})} = \frac{1}{2}\sum_{i,i'=1}^{n} Q_{i,i'}^{(\mathrm{lb})}(\boldsymbol{x}_i - \boldsymbol{x}_{i'})(\boldsymbol{x}_i - \boldsymbol{x}_{i'})^\top$$

$$Q_{i,i'}^{(\mathrm{lb})} = \begin{cases} W_{i,i'}(1/n - 1/n_y) & (y_i = y_{i'} = y) \\ 1/n & (y_i \neq y_{i'}) \end{cases}$$

- 規準:

$$\boldsymbol{T}_{\mathrm{LFDA}} = \operatorname*{argmax}_{\boldsymbol{T}\in\mathbb{R}^{m\times d}} \mathrm{tr}\left((\boldsymbol{T}\boldsymbol{S}^{(\mathrm{lw})}\boldsymbol{T}^\top)^{-1}\boldsymbol{T}\boldsymbol{S}^{(\mathrm{lb})}\boldsymbol{T}^\top\right)$$

実行例



- クラス内にクラスタ構造があっても，異なるクラスのデータを分離する部分空間が得られている
- 任意の次元数に次元削減できる

甲状腺疾患データの可視化

- 甲状腺疾患データ(5次元の検査データ)
- ラベル: 正常か異常
- 甲状腺異常には
  - 機能亢進(こうしん): 機能が強すぎる
  - 機能低下: 機能が弱すぎる

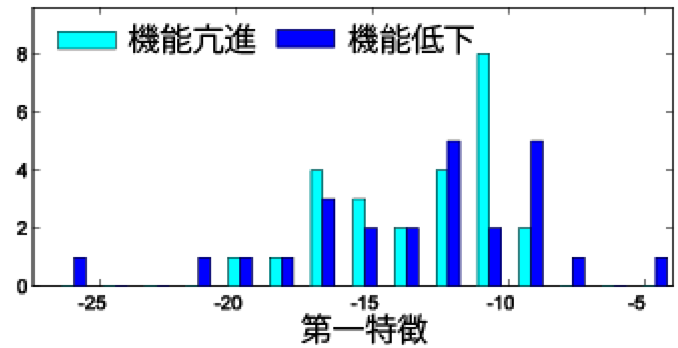
  の2種類がある

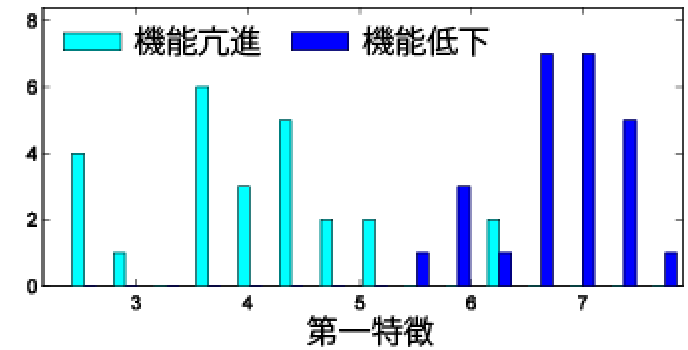
可視化結果（一次元）

フィッシャー判別分析

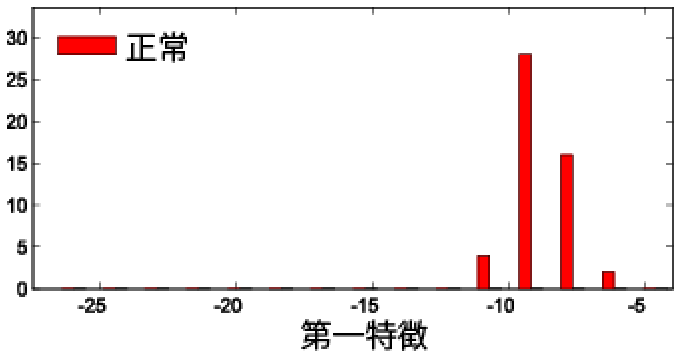
局所フィッシャー判別分析

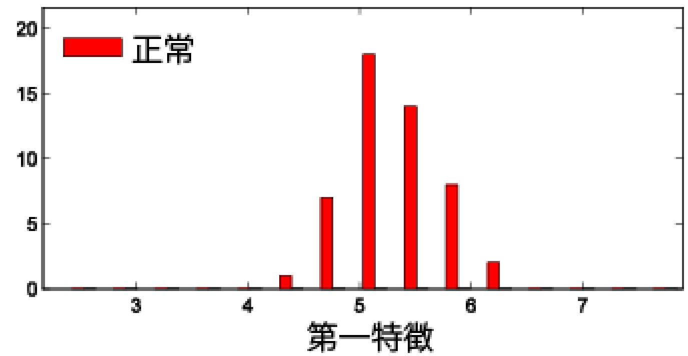
異常





正常

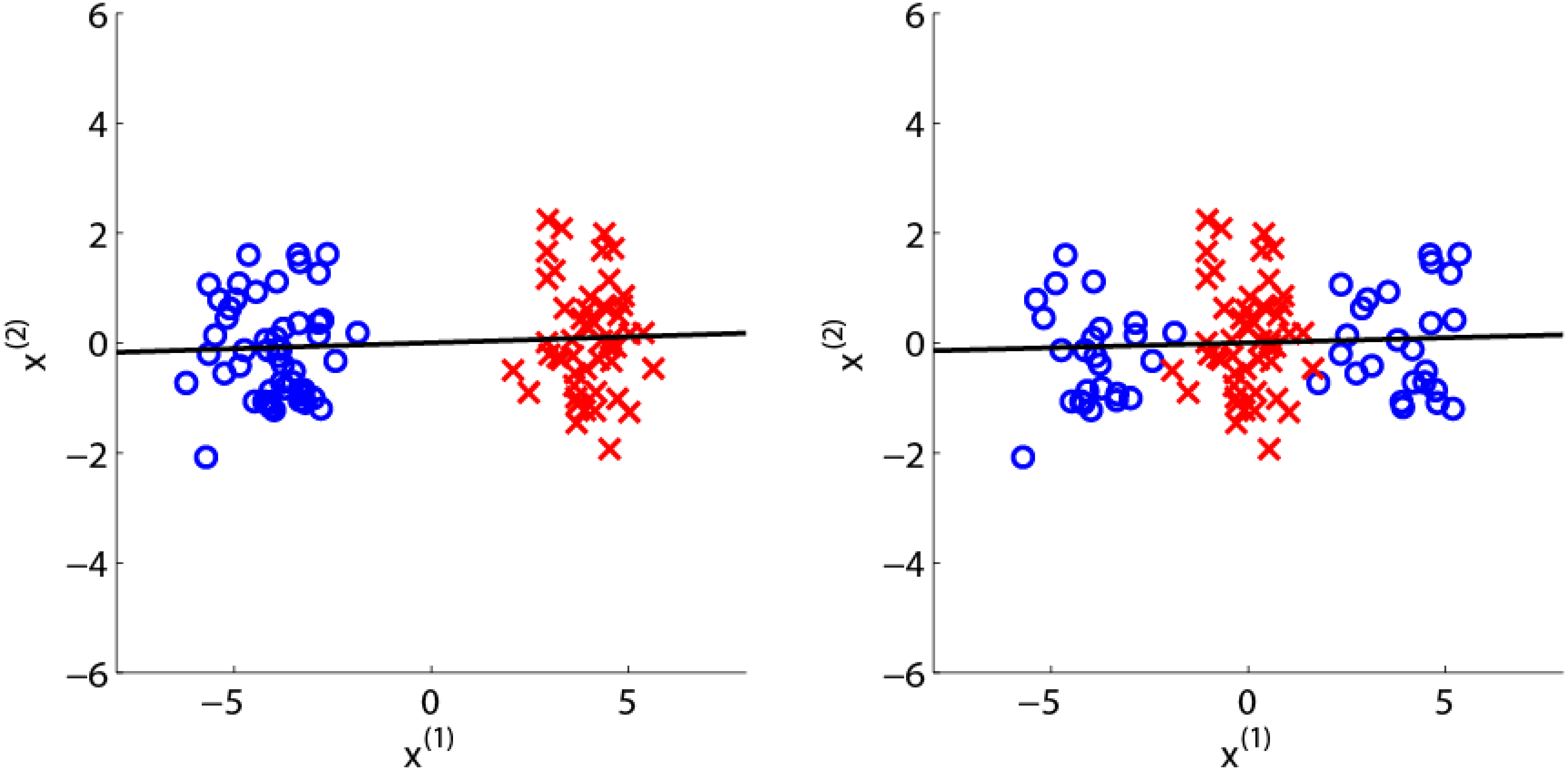




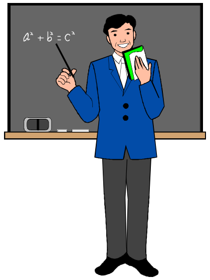
- 正常と異常はうまく分かれる
- 機能亢進と低下は混ざってしまう

- 正常と異常はうまく分かれる
- 機能亢進と低下もうまく分かれる
- 見つかった特徴は甲状腺機能のレベルと強い負の相関

局所フィッシャー判別分析：まとめ



- 一般化固有ベクトルから解析的に解が求まる
- クラス内クラスタ構造に対応できる
- 埋め込み空間の次元に上限がない
- 結果が類似度行列の選び方に依存する（交差確認可）

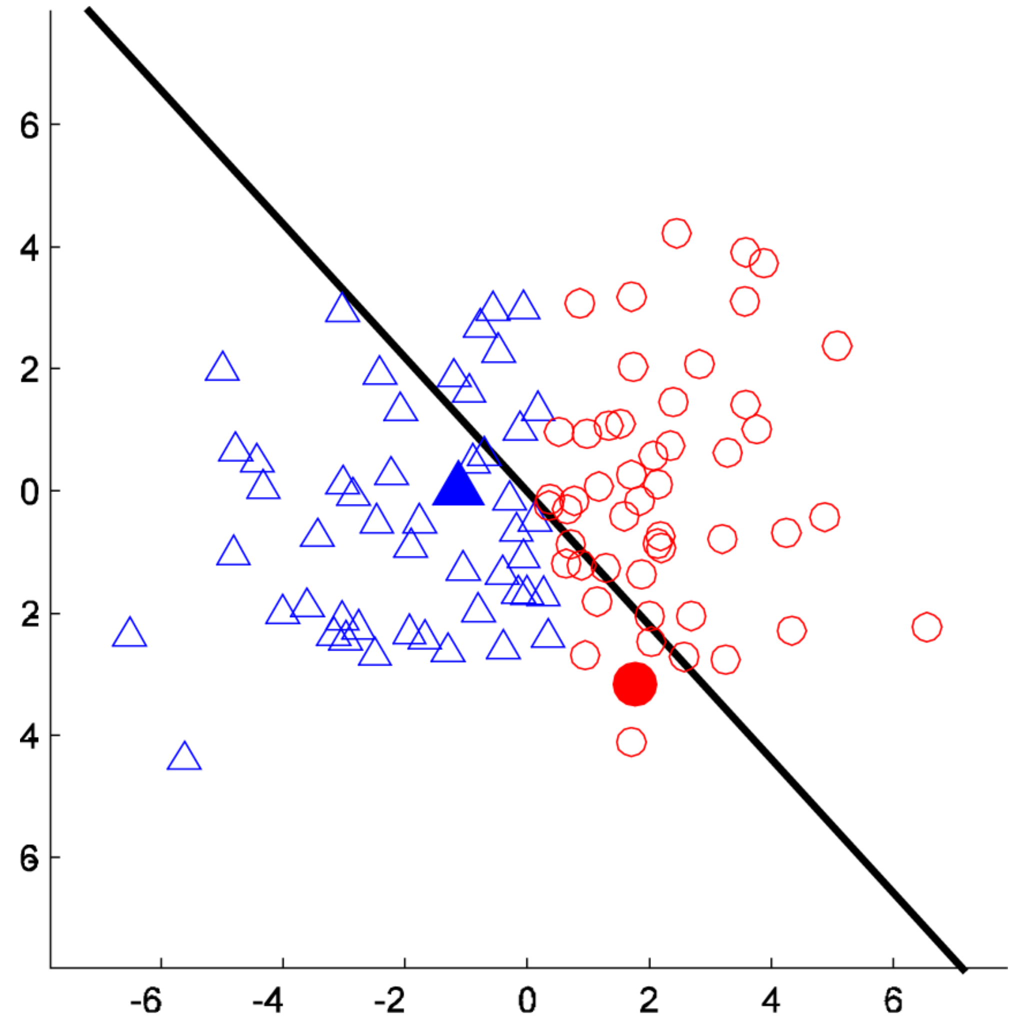


講義の流れ

1. フィッシャー判別分析（19章）
2. 局所フィッシャー判別分析（19章）
3. 半教師付き局所フィッシャー判別分析（19章）
4. 十分次元削減（19章）

教師付き次元削減の問題点

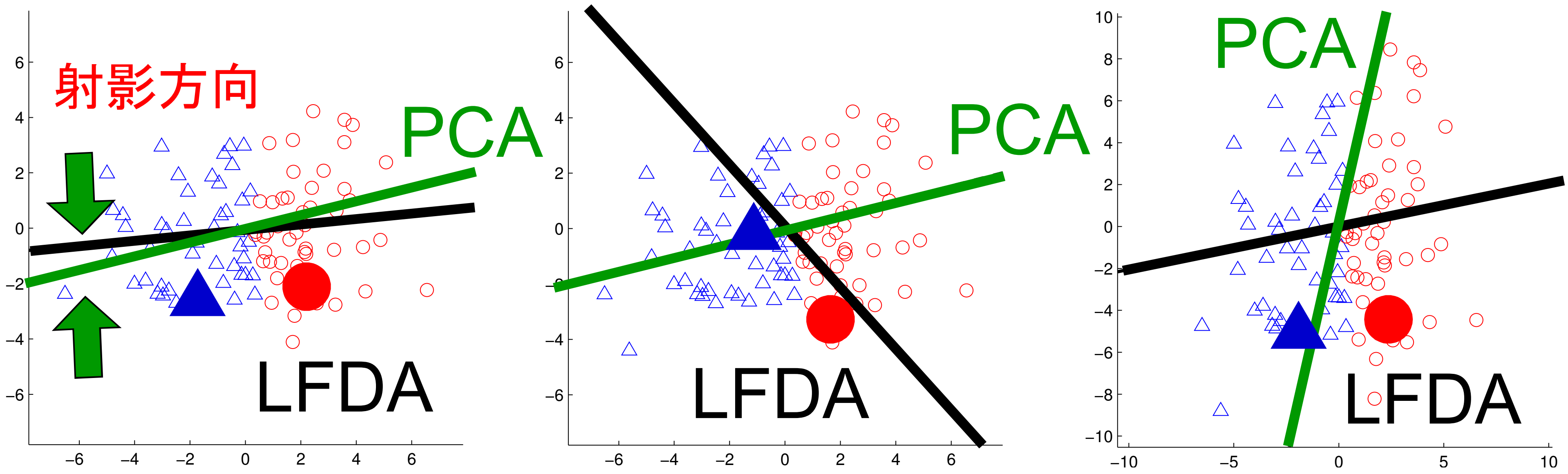
- ラベルありデータが少ない場合，過適合しやすい



半教師付き次元削減

- 半教師付き学習:
  - 少数のラベルあり標本　$\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$
  - 多数のラベルなし標本　$\{\boldsymbol{x}_i\}_{i=n+1}^{n'}$
- 教師付き次元削減法は，少数のラベルあり標本に過適合しやすい
- 多数のラベルなし標本の情報も活用したい
  - 主成分分析(PCA)は元々ラベルの情報を利用しない

半教師付き学習における LFDAとPCA



- LFDA:過適合しやすい
- PCA:ラベルの情報を利用していない
- LFDAとPCAは相補的な傾向がある

半教師付き局所フィッシャー判別分析 (SELF: Semi-Supervised LFDA)

- 基本アイデア：LFDAとPCAのいいところを組み合わせる
- 着目点：LFDAとPCAは同じ形式の固有値問題
  - PCA: $C\xi = \lambda\xi$
  - LFDA: $S^{(\mathrm{lb})}\xi = \lambda S^{(\mathrm{lw})}\xi$
- 方針：固有値問題を組み合わせる

SELF(続き)

■ SELFの固有値問題: LFDAとPCAの重みつき和

$$\boldsymbol{S}^{(\mathrm{rlb})}\boldsymbol{\xi} = \lambda \boldsymbol{S}^{(\mathrm{rlw})}\boldsymbol{\xi}$$

- 正則化局所クラス内散布行列

$$\boldsymbol{S}^{(\mathrm{rlw})} = (1-\beta)\boldsymbol{S}^{(\mathrm{lw})} + \beta \boldsymbol{I}_d \qquad 0 \le \beta \le 1$$

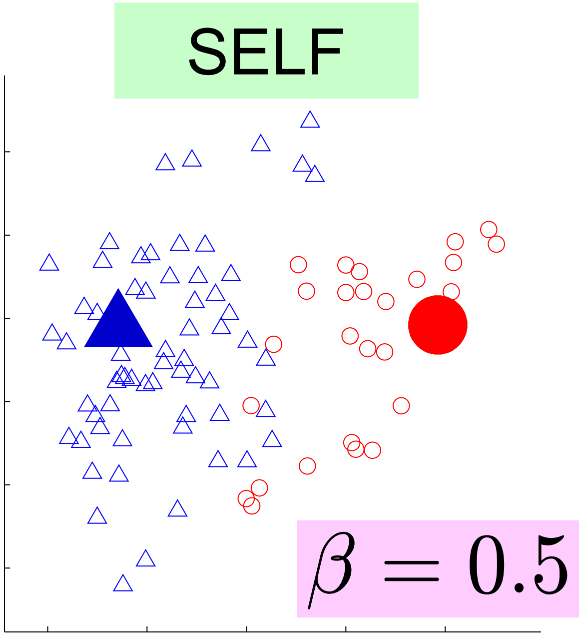
- 正則化局所クラス間散布行列

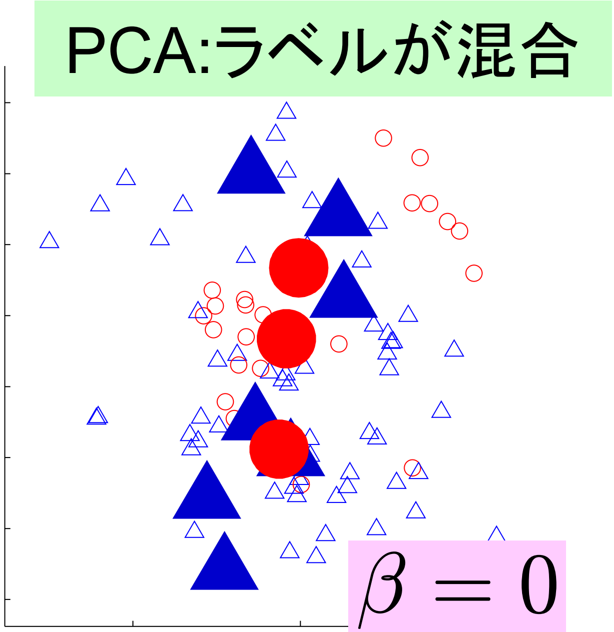
$$\boldsymbol{S}^{(\mathrm{rlb})} = (1-\beta)\boldsymbol{S}^{(\mathrm{lb})} + \beta \boldsymbol{C}$$

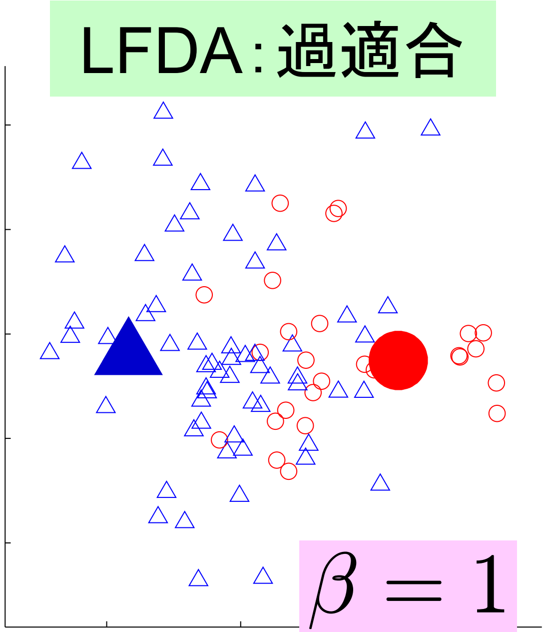
■ 解: 上位の固有ベクトルを並べる

顔画像データの可視化

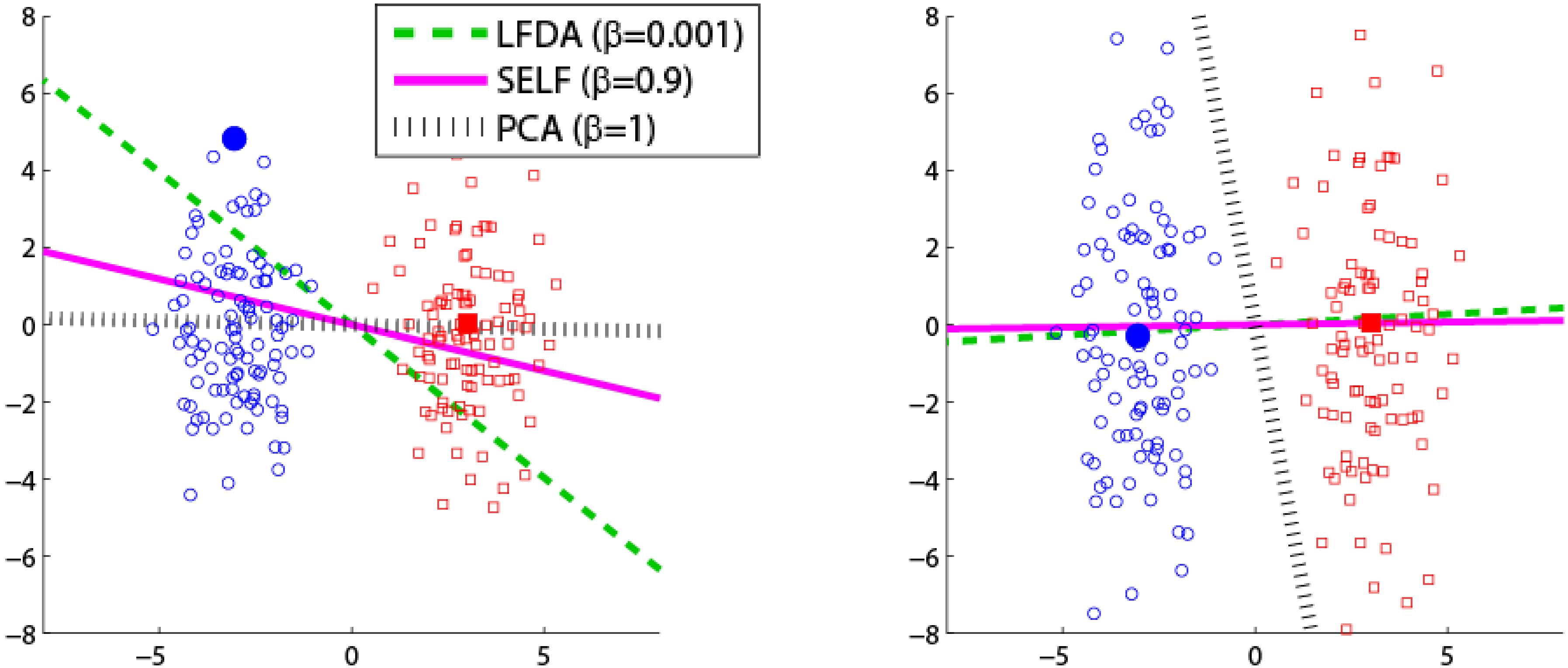
- メガネ vs. メガネなし







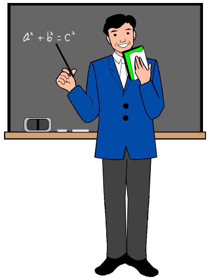
実行例



- 半教師付き局所フィッシャー判別分析によって，過適合が回避できている

半教師付き局所フィッシャー判別分析：まとめ

- 教師なし主成分分析と教師付き局所フィッシャー判別分析との組み合わせ
- 一般化固有ベクトルから解析的に解が求まる
- 過適合を回避できる
- 混合比 β の決定は難しい

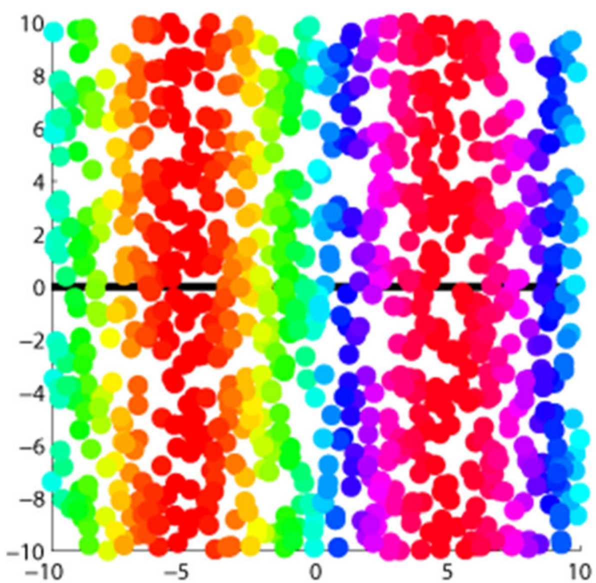


講義の流れ

1. フィッシャー判別分析（19章）
2. 局所フィッシャー判別分析（19章）
3. 半教師付き局所フィッシャー判別分析（19章）
4. 十分次元削減（19章）

十分次元削減

- 判別分析は y がカテゴリ値を取る分類専用
- y が実数値を取る回帰問題の次元削減は？
- y の情報を最大限含むzを求める
  - zを与えたとき, $\boldsymbol{x}, y$ が独立になるようにする



$$\{\boldsymbol{z}_i\}_{i=1}^n, \quad \boldsymbol{z}_i = \boldsymbol{T}\boldsymbol{x}_i \in \mathbb{R}^m$$

$$p(\boldsymbol{x}, y|\boldsymbol{z}) = p(\boldsymbol{x}|\boldsymbol{z})p(y|\boldsymbol{z})$$

（zは y の分布の十分統計量）

相互情報量

$$\mathrm{MI}(\boldsymbol{X}, Y) = \iint p(\boldsymbol{x}, y) \log \frac{p(\boldsymbol{x}, y)}{p(\boldsymbol{x})p(y)} \mathrm{d}\boldsymbol{x}\mathrm{d}y$$

- $\boldsymbol{X}$ がわかったとき Y について得られる情報量

■ 相互情報量によって独立性を測れる:

- $\mathrm{MI}(\boldsymbol{X}, Y) \geq 0$
- $\mathrm{MI}(\boldsymbol{X}, Y) = 0$ ⟺ $\boldsymbol{X}$ と Y は独立

$$p(\boldsymbol{x}, y) = p(\boldsymbol{x})p(y)$$

相互情報量

$$\mathrm{MI}(\boldsymbol{X}, Y) = \iint p(\boldsymbol{x}, y) \log \frac{p(\boldsymbol{x}, y)}{p(\boldsymbol{x})p(y)} \mathrm{d}\boldsymbol{x}\mathrm{d}y$$

$$\{\boldsymbol{z}_i\}_{i=1}^n, \quad \boldsymbol{z}_i = \boldsymbol{T}\boldsymbol{x}_i \in \mathbb{R}^m$$

- 相互情報量によって条件付き独立性も測れる:
  - $\mathrm{MI}(\boldsymbol{X}, Y) \geq \mathrm{MI}(\boldsymbol{Z}, Y)$:
    $\boldsymbol{Z}$がもつYの情報は$\boldsymbol{X}$がもつYの情報より少ない
  - $\mathrm{MI}(\boldsymbol{X}, Y) = \mathrm{MI}(\boldsymbol{Z}, Y)$

$\Longleftrightarrow$ $\boldsymbol{x}$ と y は条件付き独立

$$p(\boldsymbol{x}, y|\boldsymbol{z}) = p(\boldsymbol{x}|\boldsymbol{z})p(y|\boldsymbol{z})$$

相互情報量最大化による十分次元削減

- $$\max_{\boldsymbol{T}\in\mathbb{R}^{m\times d}} \mathrm{MI}(\boldsymbol{Z},Y) \ \ \text{subject to } \boldsymbol{T}\boldsymbol{T}^\top = \boldsymbol{I}_m$$

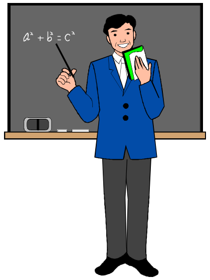
$$\mathrm{MI}(\boldsymbol{Z},Y) = \iint p(\boldsymbol{z},y)\log\frac{p(\boldsymbol{z},y)}{p(\boldsymbol{z})p(y)}\mathrm{d}\boldsymbol{z}\mathrm{d}y$$

$$\{\boldsymbol{z}_i\}_{i=1}^n, \quad \boldsymbol{z}_i = \boldsymbol{T}\boldsymbol{x}_i \in \mathbb{R}^m$$

- 勾配法により行列 $\boldsymbol{T}$ を最適化する
  - $\boldsymbol{T} \longleftarrow \boldsymbol{T} + \varepsilon\nabla\widehat{\mathrm{MI}}(\boldsymbol{T}\boldsymbol{X},Y)$ 　 $\varepsilon > 0$:ステップ幅
  - $\boldsymbol{T}\boldsymbol{T}^\top = \boldsymbol{I}_m$ を満たすように正規直交化
- しかし，一般に局所解しか見つけられない

十分次元削減:まとめ

- 出力の情報を最大限含む部分空間を探索
- 回帰, 分類両方に適用可能
- 情報量の推定に工夫が必要
- 部分空間の探索にも工夫が必要
- 実装に比較的手間がかかる

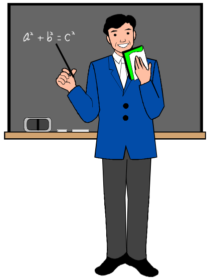


講義の流れ

1. フィッシャー判別分析（19章）
2. 局所フィッシャー判別分析（19章）
3. 半教師付き局所フィッシャー判別分析（19章）
4. 十分次元削減（19章）

まとめ

- 次元の呪い：高次元データは非常に扱いにくい
- 次元削減：データの特徴を残しつつ次元を削減
- 教師なし次元削減
  - 主成分分析(PCA)：もとのデータを保存
  - 局所性保存射影(LPP)：クラスタ構造を保存
- 教師付き次元削減
  - フィッシャー判別分析(FDA)：クラス間の分離性を保存
  - 局所FDA(LFDA)：FDAとLPPの組み合わせ
  - 半教師付きLFDA：LFDAとPCAの組み合わせ
  - 十分次元削減：回帰にも適用可能



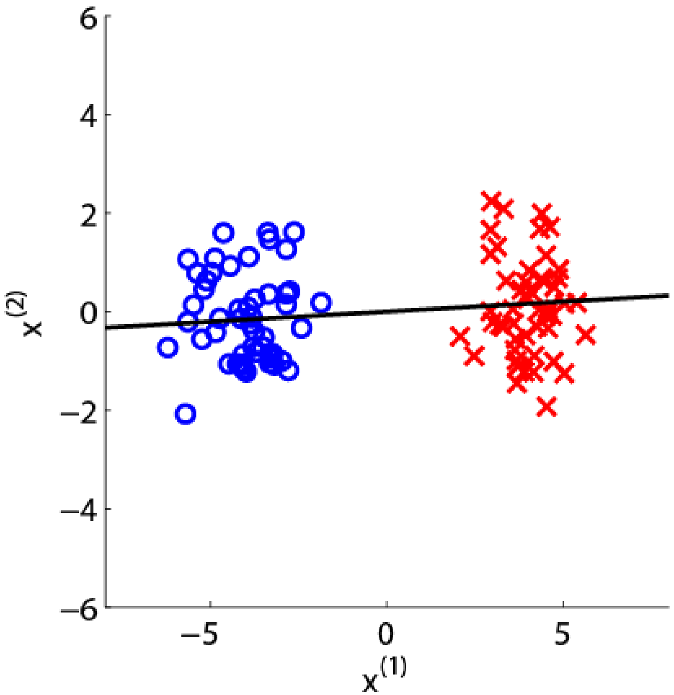
次回の予告

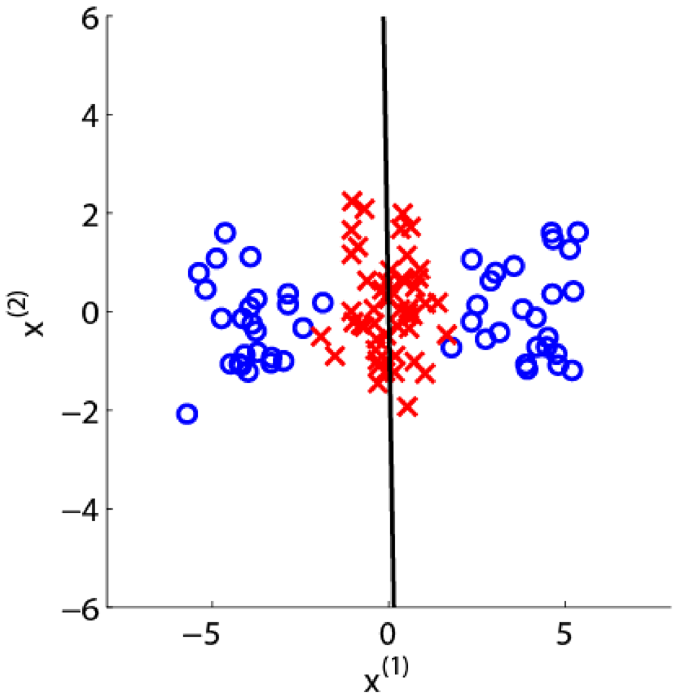
- カーネルを用いた非線形次元削減（13章）
- 来週は休講，次回は7/4です

宿題1

■フィッシャー判別分析を実装せよ:

```
clear all; rand('state',0); randn('state',0);
n=100; x=randn(n,2);
x(1:n/2,1)=x(1:n/2,1)-4; x(n/2+1:end,1)=x(n/2+1:end,1)+4;
%x(1:n/4,1)=x(1:n/4,1)-4; x(n/4+1:n/2,1)=x(n/4+1:n/2,1)+4;
x=x-repmat(mean(x),[n,1]); y=[ones(n/2,1); 2*ones(n/2,1)];
```





宿題2

- 以下の散布行列のペア表現を求めよ

n : 全標本数

n_y : クラス y の標本数

- クラス内:

$$\boldsymbol{S}^{(\mathrm{w})} = \frac{1}{2}\sum_{i,i'=1}^{n} Q_{i,i'}^{(\mathrm{w})}(\boldsymbol{x}_i - \boldsymbol{x}_{i'})(\boldsymbol{x}_i - \boldsymbol{x}_{i'})^\top$$

$$Q_{i,i'}^{(\mathrm{w})} = \begin{cases} 1/n_y & (y_i = y_{i'} = y) \\ 0 & (y_i \neq y_{i'}) \end{cases}$$

- クラス間:

$$\boldsymbol{S}^{(\mathrm{b})} = \frac{1}{2}\sum_{i,i'=1}^{n} Q_{i,i'}^{(\mathrm{b})}(\boldsymbol{x}_i - \boldsymbol{x}_{i'})(\boldsymbol{x}_i - \boldsymbol{x}_{i'})^\top$$

$$Q_{i,i'}^{(\mathrm{b})} = \begin{cases} 1/n - 1/n_y & (y_i = y_{i'} = y) \\ 1/n & (y_i \neq y_{i'}) \end{cases}$$

宿題2(続き):ヒント

$\boldsymbol{S}^{(\mathrm{w})}$ の表現:

- $$\boldsymbol{S}^{(\mathrm{w})} = \sum_{y=1}^{c} \sum_{i:y_i=y} (\boldsymbol{x}_i - \boldsymbol{\mu}_y)(\boldsymbol{x}_i - \boldsymbol{\mu}_y)^\top$$

$$\boldsymbol{\mu}_y = \frac{1}{n_y} \sum_{i':y_{i'}=y} \boldsymbol{x}_{i'}$$

$\boldsymbol{S}^{(\mathrm{b})}$ の表現:

- $\boldsymbol{C}$ のペア表現を求め, $\boldsymbol{S}^{(\mathrm{b})} = \boldsymbol{C} - \boldsymbol{S}^{(\mathrm{w})}$ を利用
- 中心化の仮定 $\sum_{i} \boldsymbol{x}_{i'} = \boldsymbol{0}$ より $\sum_{i=1}^{n} \boldsymbol{x}_i \sum_{i'=1}^{n} \boldsymbol{x}_{i'}^\top = \boldsymbol{0}$